CTAO Science Data Challenge

Fabio Pintore, INAF/IASF Palermo

on behalf of the SDC Technical Task-force Group (TTG)

CTAO Science Data Challenge (SDC)

1) it will model sources and large-scale structure finding/characterisating data challenges on simulated science-ready (DL3) data products

2) it will simulate the VHE sky as seen by the CTAO Alpha Configuration over a seven-year temporal period

3) it will allow the broad science community to get familiar with CTAO data products and the CTAO Science Analysis Tools (SATs)

4) it will serve as test bed for driving forward new algorithms and new technologies (like machine-learning) for source and large-scale structure detection/identification in the context of the source confusion

5) it will test intermediate steps in the verification process of software packages that will be used during Observatory operations and data models and formats

CTAO SDCs phases

Each SDC consists of 4 phases:

Phase 1: definition

- definition of goals, science cases and technical needs

Phase 2: preparation

- preparation and running of the simulations

Phase 3: execution

- starts with the opening of the challenge
- users can download and explore the data
- users can submit their results

Phase 4: closing-out

- starts when the challenge is closed
- it foresees the scrutiny and score of the submitted results
- it includes the writing-up of a closing-out document (peer-reviewed paper under discussion)

The first CTAO SDC - nutshell and goals

SDC#1 in short:

- blind challenge
- simulated data and SATs will be publicly released

Science-related goals (at proposal level):

- to foster preparation of blind searches for periodic emissions
- to challenge the community to detect dark matter
- to challenge the community to characterise and model the diffuse gamma-ray emission

Goals related to the CTAO Construction Project that will help:

- moving ahead tasks/deliverables that were stuck for long time
- engaging more with in-kind contributors at DPPS and SUSS levels
- producing good documentation

CTAO SDC1 – Status

SDC1 plan & definition:

- phase 1 & 2 at good point although still under discussion
- phase 3 & 4 postponed to the next months

Delivery science info: Jun 30, 2022
Execution: Jun 1, 2023 till Dec 14, 2023

Scientific deliverables: observational projects & sky models

- CTAC responsibility
- ownership of these models remains within CTAC
- hopefully Consortium papers will be submitted before SDC1 phase3 starts
- well on-track

Simulation preparation technical part:

- clear plan for IRFs
- progress in storing sky models guaranteeing the blindness as much as possible
- many points still open, but not far away from convergence (scheduling part, data quality checks)

Preparation of challenge execution:

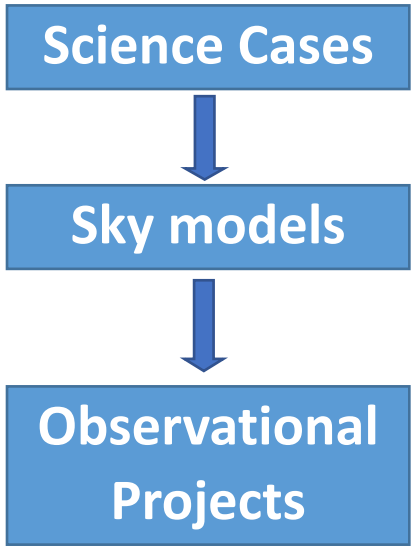
- under definition

Observational Projects

Simulated period: **seven-year** (January 1st, 2028 - December 31st, 2034).
We expect to simulate about 1200 hours of observation per year per site (TBC).

From Science Cases to Observational Projects:

1. SDC steering committee will define the ***Science Cases*** for the SDC
2. From the Science Cases we derived the ***Sky Models*** needed which will be provided by the CTAC to the SDC TTG
3. The Sky models can be grouped in ***Observational Projects.***



Observational Projects

The observational projects can be subdivided in the following main core projects.

Four of them are included in the **sky surveys**:

- Galactic plane
- extra-galactic sky
- Large Magellanic cloud
- Perseus cluster

while the remaining are mostly connected to **variable and exotic sources**, namely:

- AGN monitoring (15 AGNs)
- ToO dedicated to follow up of neutrino events, Galactic transients and MWL transient sources,..
- GRBs (rapid response)
- GW follow up (pointing scan)
- Primordial black holes
- other dark matter sources

In addition, possible OFF observations may be provided for tests.

Sky models – Galactic sources

Supernova remnants and pulsar wind nebulae (GPS Consortium paper)
Based on models from Cristofari et al. (2013,2017) & Fiori et al. (2021)

Binary systems (GPS Consortium paper)
Based on Dubus et al. (2017) with spatial distribution of OB stars and and spin-down power distribution of pulsars

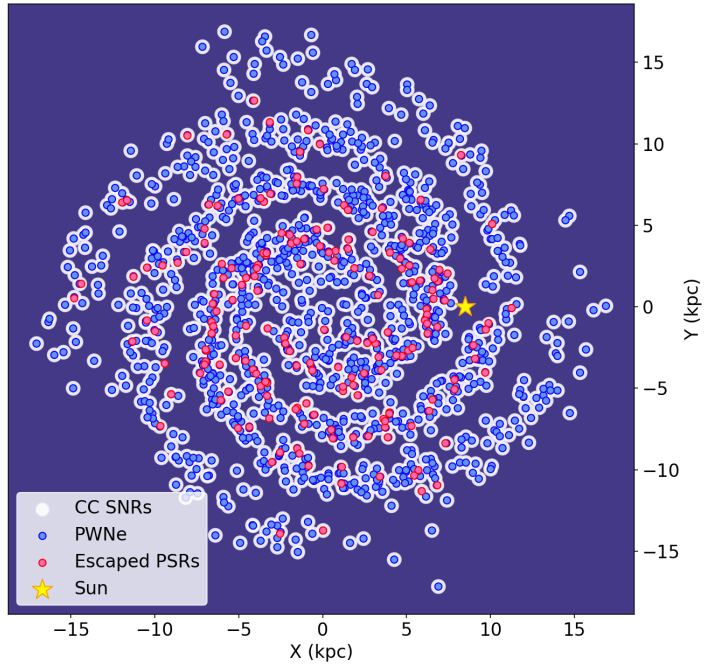
Stellar clusters (star-forming systems consortium paper)
Cosmic rays (CR) production based on acceleration at cluster wind termination shock from Morlino et al. (2021)

Pulsars (GPS Consortium paper)
The previous DC based on 3FHL pulsars (phase-averaged spectra, with possibly PWN pollution)
→ now we will use Fermi 2PC

Ultra-high-energy sources/PeVatrons (GPS Consortium paper)
Driven by LHAASO detection

Galactic transients
Transient low mass XRB, Cyg X-1, SS433, GX339-4

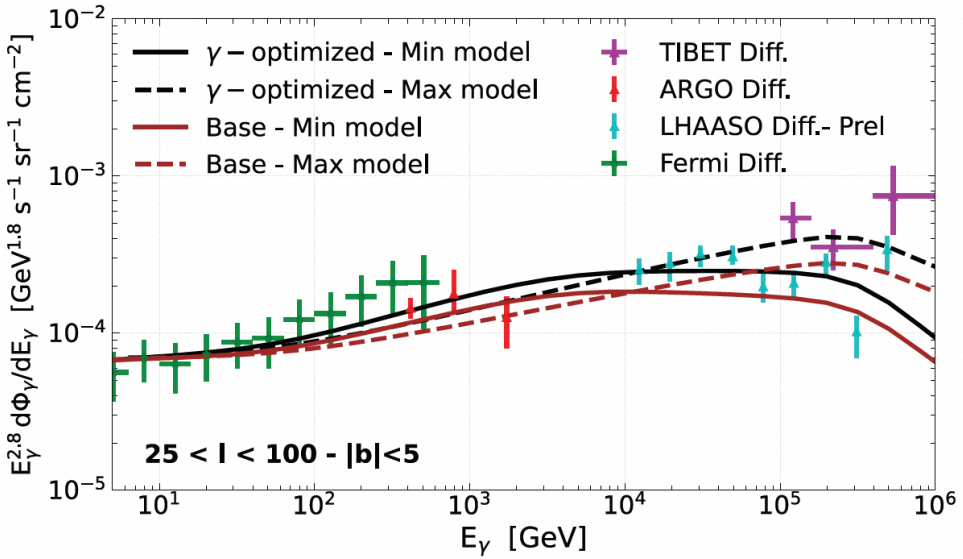
Need to merge synthetic populations and known objects to avoid “double-counting”



Sky models – Diffuse and LMC

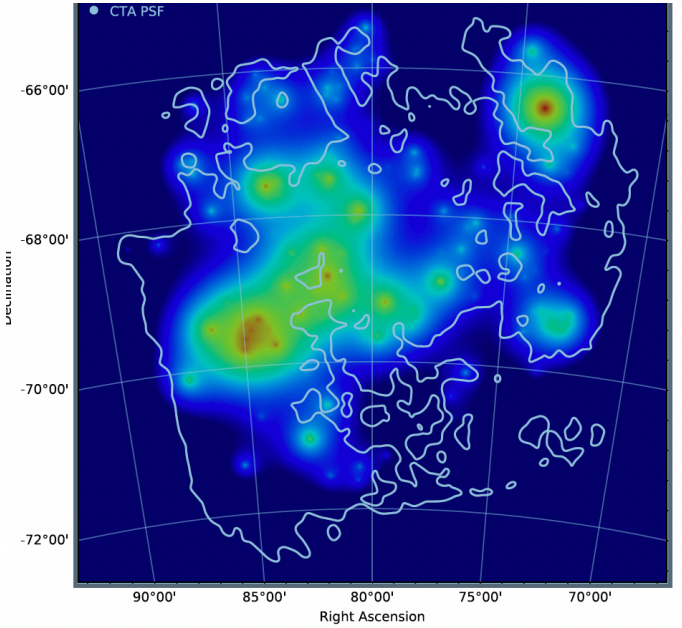
Galactic diffuse emission (GPS Consortium paper)

- Interstellar models from De la Torre et al. (2022)
- Analytical model for Fermi bubbles extrapolated from LAT data, possibly with energy-dependent morphology



LMC (LMC Consortium paper)

- Similar to Milky Way
- Interstellar emission: particle injection following HII regions plus transport in average interstellar conditions



Sky models – EGAL and AGN flares

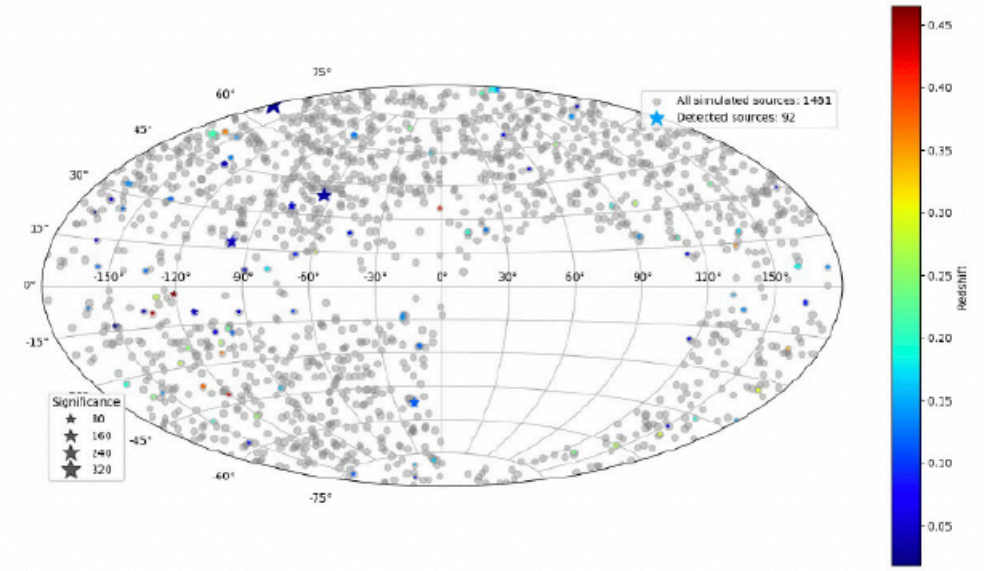
AGN population: 4LAC sources
Extrapolation of Fermi-LAT 4LAC AGNs with revised redshift (P. Goldoni)

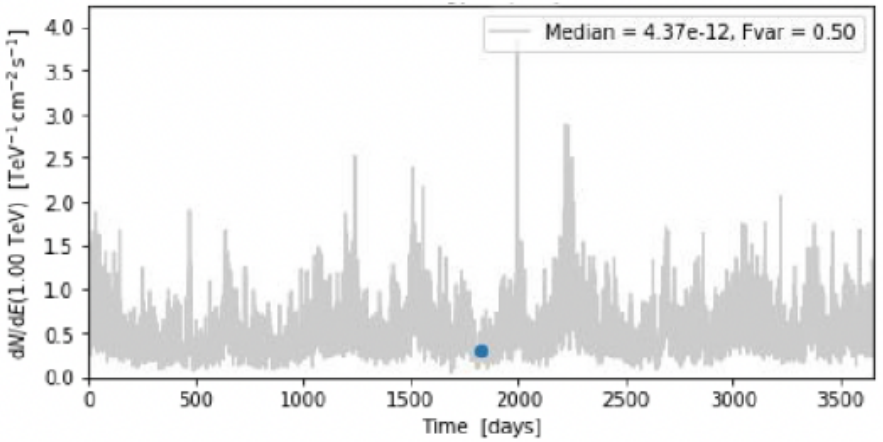
AGN variability for long term monitoring (16 AGNs)
to be monitored 30 min/week
2 MWL campaigns

AGN flares
Flares (few hours) from few occurency (Mrk 421, BL Lac, 3c279..)

Pair halos
one instance of IGMF-induced pair halo

Starburst galaxies
Model based on Peretti et al. (2019) for particular sources (M82, NGC253, Apr220)





Sky models – GRBs, GW and neutrinos

GRB

select few GRBs from the present Long Afterglow simulation

GW follow up

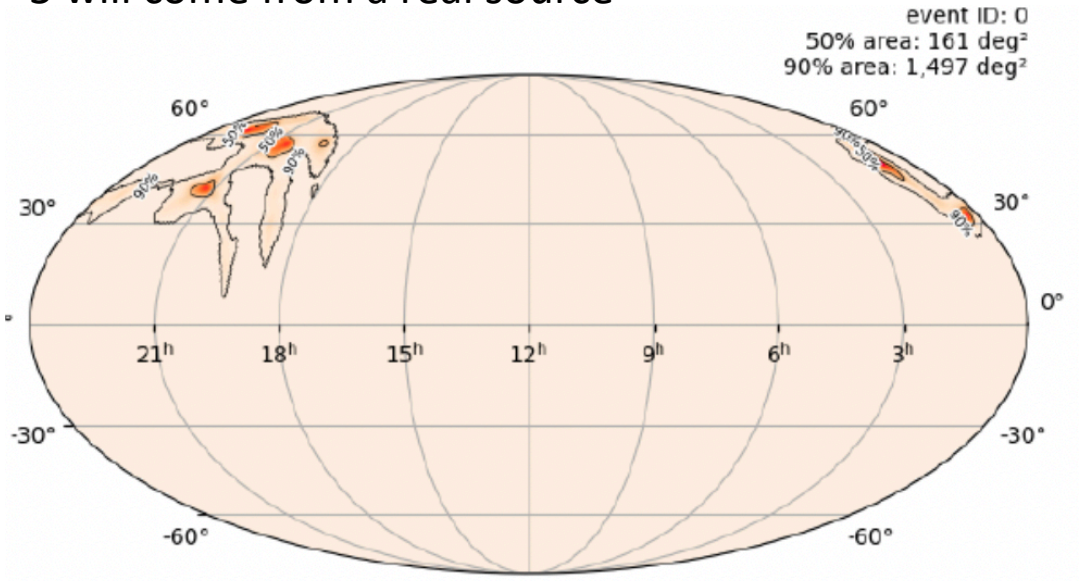
simulations of BH-NS mergers

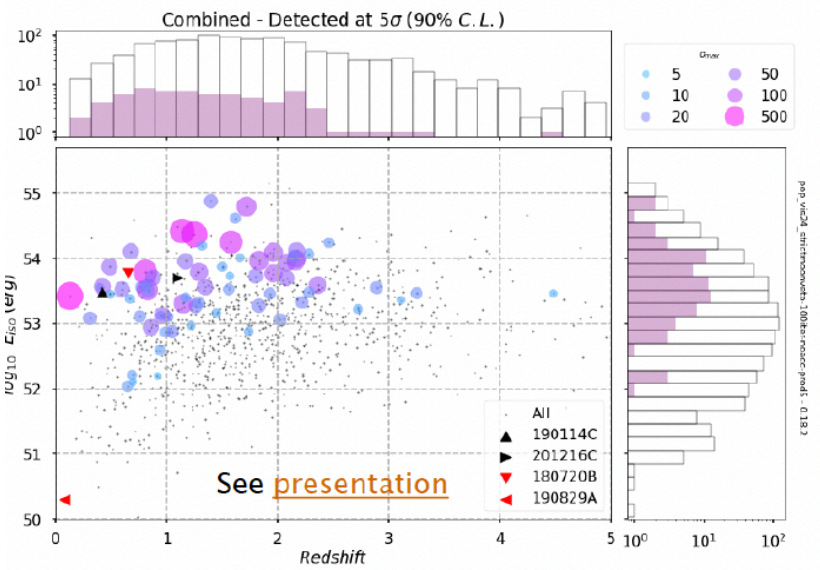
Synthetic GW and GRBs spectra and lightcurve

Neutrino sources

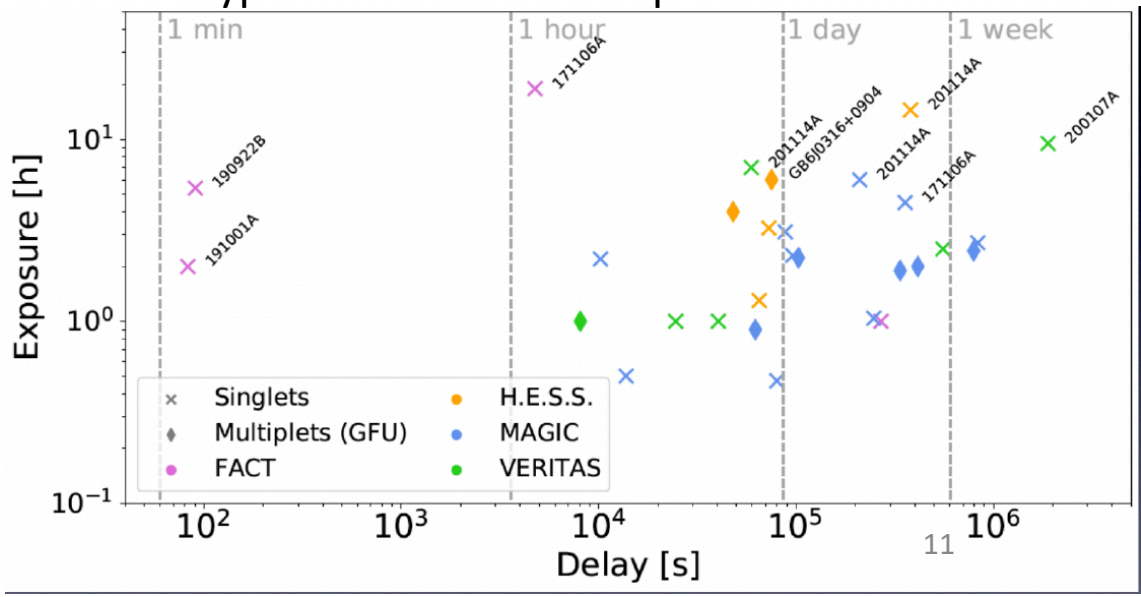
Out of 10 Gold alerts/year;

5 will come from a real source





Typical current IACT response on neutrinos



Sky models – Exotic sources

Galaxy clusters (Perseus Consortium paper++)

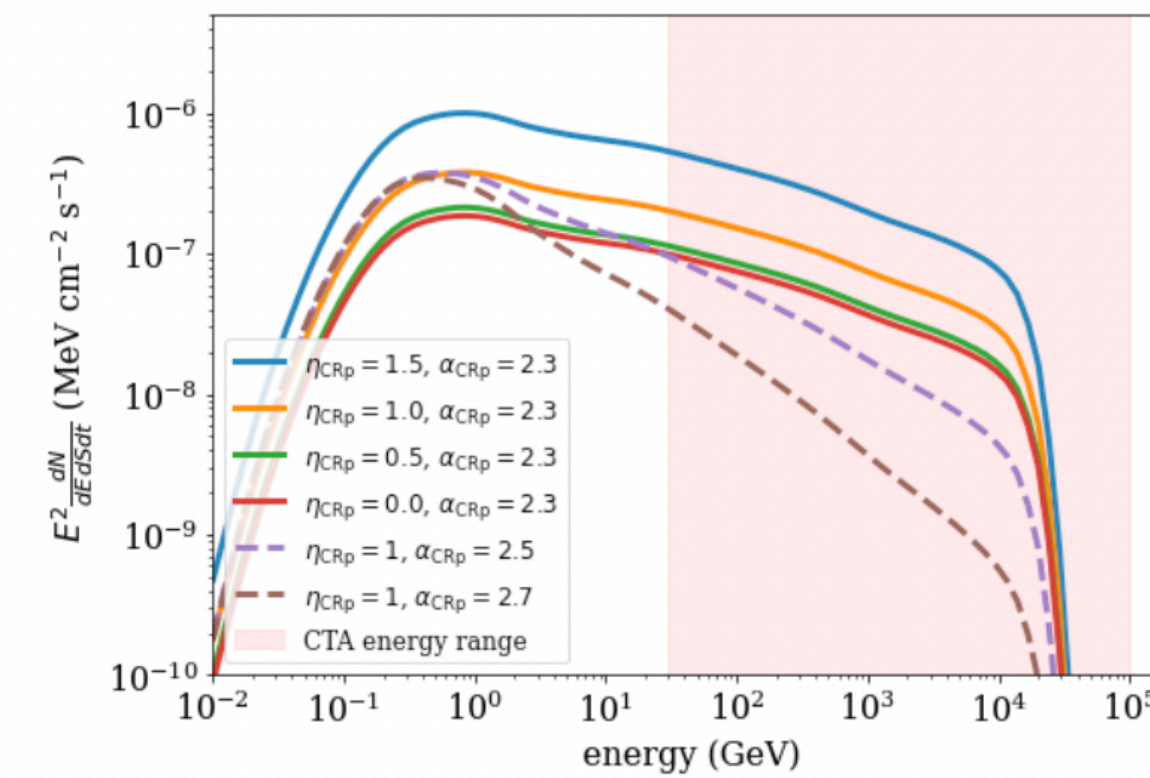
- Perseus cluster: gamma-ray emission dominated by hadronic interactions
- All-sky model from X-ray catalogs based on mass and redshift, plus self-similarity

WIMP dark matter (DMEP SWG consortium papers++)

- Benchmark for various DM masses/cross-sections/ annihilation channels (particles)
- Astrophysical motivated models for DM distribution in: Galactic center, known dwarf galaxies, LMC, Perseus cluster, Dark sub-halos (at the positions of some Fermi unidentified sources)
- Some scenarios will include detectable dark-matter signals

Primordial black holes

- Bursts from the last seconds of the PBH life



Observational Project templates (1)

cherenkov telescope array

Observational project template for the Science Data Challenge

1 General section

- **Project name:**
  Please, indicate also the email address.

- **Contact person:**

2 Scientific section

- **Scientific Rationale** [max ½ page]

- **Science Targets** [celestial equatorial coordinates]

Note: a science target is a celestial object or region of interest that is to be observed for executing a scientific proposal (e.g., for a survey the science target is the whole sky region to be surveyed).
In case of ToO simply say that they are ToO.

Observational Project templates (2)

3 Technical section

- **Telescope tracking mode: sideral mode**
  Note: no other option is available during the SDC#1
- **Array Pointing mode: parallel pointing mode**
  Note: no other option is available during the SDC#1
- **Observation mode:**
  - ☐ **wobble**
    - **number of wobbling positions: 5**
      Note: no other option is available during the SDC#1
    - **offset angle: 0.7°**
      Note: no other option is available during the SDC#1
    - **offset rotation angle: 0°**
      Note: no other option is available during the SDC#1
  - ☐ **on/off**
    Note: the dark sky region (off) is scheduled by the SDC team
  - ☐ **scan (user-defined grid)**
    Please, describe below the pointing grid you are requesting as a list of pointing positions and include the corresponding algorithm, if available. The latter case for instance is the double row of the GPS. The pointing positions should be provided in machine-readable format. Therefore, indicate here simply the name of the file containing those positions.
- **Total requested observing time [hours]:**
- **Requested observing time per SB target [hours]:**
  Note: this figure applies also to ToO observations
- **Temporal requirements on the SB targets:**
  Please, describe below the algorithm that specify the cadence if relevant for your project (i.e. periodic emission) or the follow-up strategy in case of a ToO, otherwise simply let the figure empty.
  Note: this figure applies also to ToO observations
- **Minimal array configuration:**
  - **Northern Array Alpha Configuration**
  - **Southern Array Alpha Configuration**
- **Zenith distance range**
  - ☐ **20°**
  - ☐ **40°**
  - ☐ **60°**
- **Azimuth angle:**
  - ☐ **North only**
  - ☐ **South only**
  - ☐ **Averaged**

Observational Project templates (GW only)

cherenkov telescope array

Pointing strategy for the follow-up of a gravitational wave alert

For the gravitational wave a special proposal for the pointing strategy has been prepared.

- **Pointing Strategy**

Note: please describe the algorithm and indicate all the relevant inputs, if any (i.e. which galaxy catalogue). You may consider to attach a Jupiter notebook.

Instrument Response Function

To generate the gamma-ray simulated sky as seen by CTAO one would need to convolve the sky models with the *Instrument Response Functions* (*IRFs*) of the two considered CTAO arrays.

For the SDC we will be using the publicly-released IRFs that will result from the ongoing Monte Carlo **prod-v6.** In this production we foresee:

- three ***zenith*** distances: 20° 40° 60° (we are trying to consider a further bin at 50°)
- two ***azimuth*** angles: 0°, 180° (corresponding to pointing towards North and towards South)
- two ***NSB*** bins corresponding to dark observation conditions and half-moon conditions.

Northern array and Southern array will be considered in their *Alpha Configuration* and final layout.

Please note that if the Italian funding application will be approved before the start of the simulation running, we can consider, at the last minute, to use the *Beta Configuration* which is defined as the Alpha one plus the telescopes funded by the above-mentioned funding application

Instrument Response Function

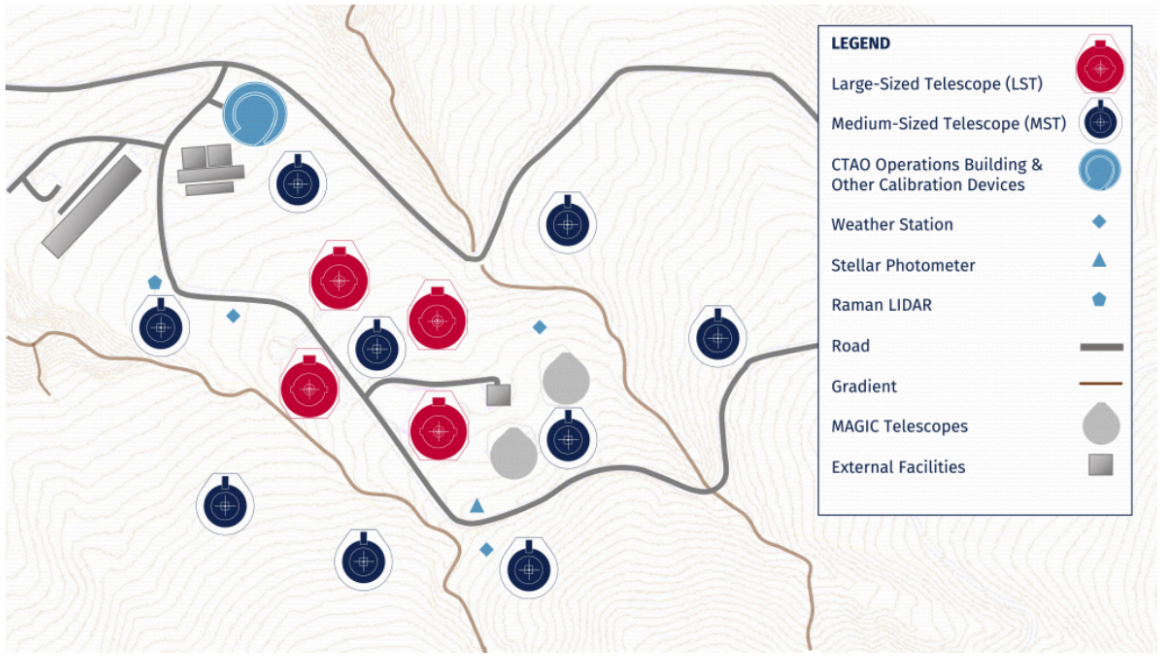


Figure 1. Layout of the CTAO Northern array on La Palma (Spain), including the elements defined in the Alpha Configuration.

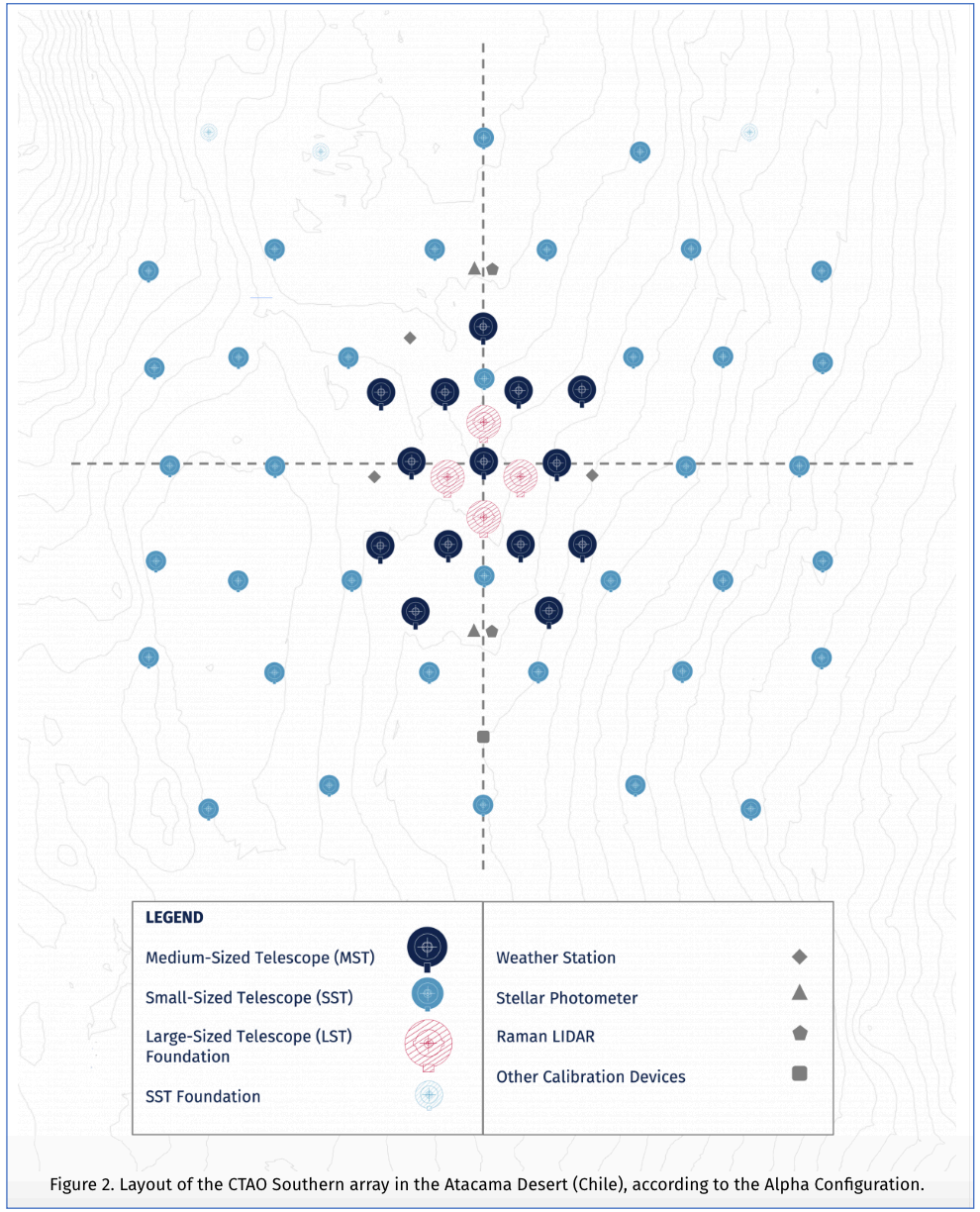


Figure 2. Layout of the CTAO Southern array in the Atacama Desert (Chile), according to the Alpha Configuration.

Sky model and obs. project templates collection

The TTG collects the sky models (created both with CTOOL-XML and Gammapy-YAML formats) and the observational project templates.

We will then convert all CTOOL models to Gammapy format (YAML file), since we will use Gammapy for the simulations.

The simulated data released by the SDC are DL3 data products

- So far no official DL3 data format has been released by CTAO (coming soon)
- For the time being while testing the simulator we are using the DL3 adopted by Gammapy (GADF)

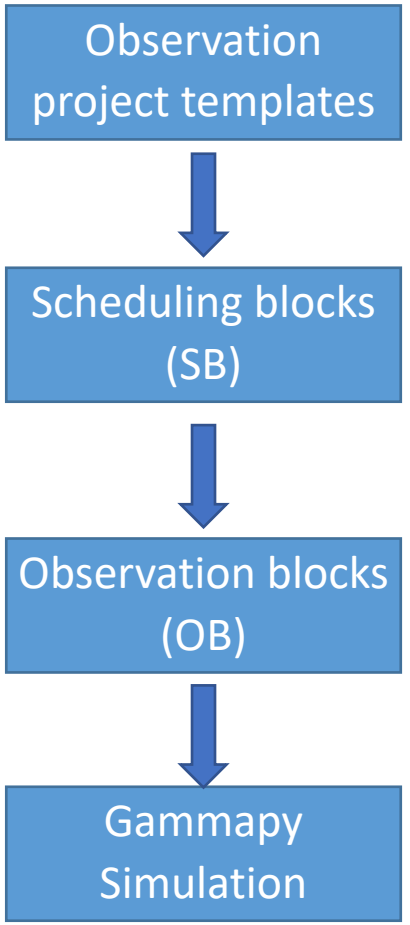
https://docs.gammapy.org/0.20/_images/data-flow-gammapy.png

SDC construction project - scheduler

1. Starting from the Observational Project Templates we plan to generate a mock catalogue of Scheduling Blocks (SB) following the SB data format already released by CTAO.

2. From the SB an “Observation Block” (OB) will be created. It is currently under discussion how to generate the schedule and the corresponding OB list. The schedule generation will take into account the following hard constraints:
   - Source visibility
   - Zenith distance
   - Fixed Time Window → the way we will handle ToO observations
   - Observing time request
   - Pointing cadence
   - No sub-array available

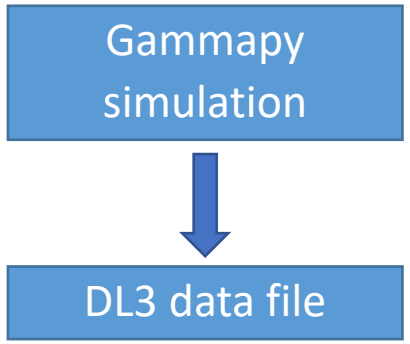
3. The input for the simulator (part of the SAT) is the list of OB. The conversion from OB data format to Gammapy input format will be discussed in detail with the Gammapy team or otherwise handled by the TTG.



SDC output

GADF format relies on several FITS HDUs:

- EVENTS: table of gamma-like events measured parameters
- GTI: Interval of time associated to events
- POINTING: Telescope pointing info
- AEFF : Effective area table (true energy, FoV offset)
- EDISP : Energy dispersion (true energy, FoV offset)
- PSF: isotropic PSF (true energy, FoV offset)
- BACKGROUND: (energy, FoV lon, FoV lat)
- A general HDU table connects everything



For more information look at:
https://docs.gammapy.org
https://github.com/AtreyeeS/VHE-KU-workshop/blob/main/data_exploration.ipynb

Limitations of the simulated data products

While significant effort has been expended to make a **realistic** data product for the SDC1 analysis, there are many limitations to the degree of realism that could be achieved.

Some of the *most apparent limitations* are :

- assumption of brand-new array elements
  -> systematic errors due to telescopes getting older are not considered,
- assumption of excellent weather conditions
  -> systematic errors due to the changing atmospheric conditions are not considered,
- no observations with different sub-arrays
  -> only observing in parallel at the same site is considered,
- The zenith distance and azimuth angle binning used for the IRFs production is rather coarse
  -> this will cause significant deviations from reality

This, however will be the 1st of a series of SDCs therefore we may expect many improvements in the future!!

Execution preparation - ESCAPE tech

ESCAPE (https://projectescape.eu/)
European Science Cluster of Astronomy & Particle physics ESFRI

- Group of ESFRIs working on common solutions to commons problems
- A&A system, Data lake, software repository, links to IVOA & Science Platform

SKAO Science Data Challenges

- One lesson learnt: important to be able to reproduce the results of participants
- Control the environment: participants able to run workflows on provided systems

Findable, Accessible, Interoperable and Reusable (FAIR) principles:

- will help us to understand the results
- unique and persistent identifier
- metadata
- detailed provenance information
- data and software licences

Outlook

All the people involved (CTAC and CTAO) are actively working on the preparation of the first CTAO SDC!

- Start of the execution of the **SDC** expected in **mid 2023**.
- The idea is to deliver the SDC simulations in bunches (most likely of 1 year each) with a given cadence (e.g. every month)
- **Winners' nominations: Jan 2024!!**

Thank you very much for your attention!

Backup slides

CTAO SDC1 – Governance

Steering Committee (SC)

- **has the responsibility to steer the challenge, define the goals and prepare an execution plan**
- chaired by the PS
- members:
  - Matthias Fuessling (SUSS coord)
  - Karl Kosack (DPPS coord)
  - Alba Fernandez (Communication Officer)
  - Giacomo Principe (chair of the TTG)
  - Luigi & Franz (CTAC Science coordinators)
  - Regis Terrier (gammapy lead developer)

In addition there are:

- Technical Task-force Group (TTG) which has the responsibility to guarantee a smooth execution of the challenge.

- SDC Advisory Committe (SDCAC) which has the responsibility to advice the SC